# Exemplo MapReduce - Wordcount

<https://campusformacionnebrija.imf.com/general/prt/prg/leccgen/newud.php?codcurso=C7274C13&unidad=2&trozo=18&param=contenido>

Vamos a crear un directorio en el que después crearemos un par de ficheros de texto sobre los que después trabajaremos.

*$ sudo mkdir wc-in*

*$ sudo chmod -R 777 wc-in*

*$ sudo echo "bla bla" > wc-in/a.txt*

*$ sudo echo "bla wa wa" > wc-in/b.txt*

Copiamos los ficheros anteriores a un directorio en HDFS. Esta vez emplearemos el comando copyFromLocal que es similar al comando put con la salvedad de que el fichero tiene que tener una referencia local.

*$ hdfs dfs -mkdir /user/bigdata/dir4*

*$ hdfs dfs -copyFromLocal wc-in /user/bigdata/dir4*

Vamos a emplear un ejemplo de MapReduce para **saber el número de palabras que contienen los ficheros creados anteriormente**.

*$ hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.0.jar wordcount /user/bigdata/dir4/wc-in /user/bigdata/dir4/wc-out*

Los resultados pueden verse a través del sistema de archivos:

*$ hdfs dfs -cat /user/bigdata/dir4/wc-out/\**

*bla 3*

*wa 2*

Para archivos grandes, no se pueden emplear los comandos -more, -head o -tail. En su lugar hay que usar la tubería (pipe “|” ) junto con el comando cat.

*hdfs dfs -cat /user/bigdata/dir4/wc-out/\* | more*